

DOCUMENT RESUME

ED 233 070

TM 830 519

AUTHOR Kraemer, Helena C.aura
TITLE Unorthodox Uses of Correlation Coefficients.
PUB DATE 11 Apr 83
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; Behavioral Sciences; *Correlation; *Research Methodology; *Statistical Analysis
IDENTIFIERS Data Interpretation; Effect Size; F Test

ABSTRACT

While the potential use of correlation methods has spread to wide contexts and these uses have permeated behavioral research, their scope remains the orthodox applications: to test independence, and to measure association. It is well worth noticing that problems having no ostensible connection to dependence, independence, or association have been resolved by use of correlation theory. Three examples of unorthodox uses are: (1) compare means, (2) compare variances, and (3) properly prepare for and interpret profile analysis of variance results. In place of comparing the mean difference in two independent normal populations having equal variances, one might report the effect size (the mean difference expressed in units of within-group standard deviation) approximated by a correlation coefficient. In samples of two normal distributions, if the data were totally paired, one could also test the equality of variances with a correlation coefficient. To interpret profiles from analysis of variance, a discarded F-test for items is equivalent to correlation between paired subjects' responses as a measure of the similarity of "shape" of two profiles. (CM)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED233070

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

* This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

UNORTHODOX USES OF CORRELATION COEFFICIENTS

Helena Chmura Kraemer

Department of Psychiatry and Behavioral Sciences

Stanford University

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

H. C. Kraemer

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Presented at the American Educational Research Association in Montreal,
Canada. April 11, 1983.

Introduction

Between the years 1880 and 1915, a major part of the theory and practice underlying current research use of correlation coefficients was formulated. In this original formulation, the correlation was conceived for use in one context, a bivariate normal population, and for two closely related uses.

1. As a test statistic to test independence;
2. As a measure of association interpretable both in terms of strength and direction of a linear relationship and in terms of percentage variance accounted for by a linear relationship.

Since then, the use of the correlation coefficient has spread to other contexts. If one variable is a dichotomy and the other measured on a continuum, it is used and called the Point Biserial Correlation Coefficient. If both are dichotomies, it is used and called the phi-coefficient. If both are rank-ordered, it is used and called the Spearman coefficient. If both are replaced by standard normal deviates corresponding to their percentile ranks, it is used and called the Fisher-Yates coefficient.

The correlation coefficient is also used in more specialized contexts. Used on twin data, it is a heritability coefficient. Used on two measurements of the same thing by two different observers, or two different times, or under two different conditions, or using two different instruments of measurement, it becomes a reliability coefficient. Used on one measurement and a criterion measurement of the same thing, it becomes a validity coefficient.

Since 1915, too, the use of the correlation coefficient and its associated theory has become very widespread, particularly in the behavioral sciences: education, psychology, sociology, economics and the behavioral aspects of medicine. A survey of articles in the Applied Psychology literature (Moore, 1975) reported that 67% of the research articles using any test or method at all, used some correlational method (including e.g. Factor Analysis). Moore suggests that teachers of statistical methods be more aware of which methods are most commonly used in which fields.

There is indeed a disparity between the usage of correlational methods in these research fields and the relative time invested to train both statisticians and researchers who will work in these fields, in correlation theory and application. A semester is spent on Analysis of Variance and a week or two, perhaps, on correlation theory.

Generalizations are dangerous, but a well-trained new Ph.D. in Statistics knows the sample and population definitions of correlation coefficients, and is likely to know the null distribution of the sample correlation coefficient. What she/he knows about non-null distribution may date back to 1915. She/he may be under a general impression that the distribution theory is very non-robust, not to be used with other than bivariate normal data. She/he may also share a sentiment expressed most succinctly by an anonymous reviewer: that those who really knew what they were doing would use regression, not correlation methods.

On the other hand, a well-trained new Ph.D. in behavioral research knows the definition of the sample correlation coefficient, but may have little idea why it is defined as it is. She/he will know how to test the null hypothesis of independence at least to the extent of knowing where to locate critical values, but will know little about its non-null uses. She/he will have a general impression, fostered by numerous articles in the

behavioral research literature, that the distribution theory is totally robust. (Kraemer, 1980) applicable to any or all paired data. Finally, she/he may use these methods more frequently than any other single statistical method.

The fact that we spend so little training effort on correlation is certainly not due to a lack of knowledge or of progress since 1915. Compilation of the statistical research literature dealing with correlation in any single year since 1915 is substantial. It may be more closely related to the fact that while the potential use of correlation methods spread to wider contexts, and these uses have permeated behavioral research, the scope of such use remains the orthodox ones: to test independence and to measure association. If one believes that estimating and interpreting the correlation coefficient and testing whether it is zero, is essentially all there is, why should there be any further educational investment in this area?

Therefore, it is well worth noticing that problems having no ostensible connection to dependence, independence or association can and have been resolved by use of correlation theory. To demonstrate this point I have three examples of such unorthodox uses: 1. To compare means; 2. To compare variances and 3. To properly prepare for and to interpret Profile ANOVAs.

To Compare Means

Suppose we sampled two independent normal populations having equal variances, and wished to compare their means. Of course, the obvious solution is to compute sample means, report their difference and test the null hypothesis of equality using a t-test.

The difficulty with the obvious solution is that response measures

are often uncalibrated measures. A mean difference of 10 units in one study might be quite different from a mean difference of 10 units in another. Whether a reported mean difference is to be regarded as large or small is frequently problematic. It has been recommended (e.g. Glass, 1981 and other proponents of Meta-Analysis) that in addition to or in place of the mean difference, one might report the effect size (ES), the mean difference expressed in units of within group standard deviation.

What has this to do with correlation coefficients? As it happens, the distribution of the effect size can be approximated by that of a correlation coefficient (Kraemer, 1981). Thus there is a transformation of the effect size:

$$r = ES / \sqrt{ES^2 + (N-2)/N_{pp}}$$

where N = total sample size

N_p , $N_{p'}$, the sample sizes in each group, having approximately the distribution of a correlation coefficient.

Consequently to test whether an effect size is zero, one may test whether its transformed value is zero. This is, in fact, the usual t-test. To obtain a confidence interval for the population effect size, one may obtain a confidence interval for its transformed population value and convert it. To stabilize variance, one can use Fisher's z-transformation known to stabilize variance for the correlation coefficient to stabilize variance of the transformed value. To test homogeneity, one may test the homogeneity of transformed values. In short, all the testing and estimation techniques developed over most of this century for correlation coefficients can be applied to analogous problems for effect sizes.

In fact, this approach produces results parallel to those developed by Hedges and Olkin (1981 a,b) using a different approach, but remain a good example of an unorthodox use of correlation theory.

To Compare Variances

Suppose one samples two normal distributions and wishes to compare variances. If the samples are independent, one computes the sample variances (S_x^2, S_y^2) , the variance ratio $(\tau = S_y^2 / S_x^2)$ and tests the null hypothesis using an F-test. However, what if the samples were paired, or partially paired, or grouped in some way?

In 1939, Morgan (1939) and Pitman (1939) independently pointed out that if the data were totally paired, one could test the equality of variances with a correlation coefficient. Specifically, the sum and difference of paired values $(x+y$ and $x-y)$ are uncorrelated if and only if their variances are equal. Thus if one replaces the original paired data by their sums and differences and computes the correlation coefficient between these:

$$r(\tau, \hat{\rho}) = (\hat{\tau}-1) / \sqrt{(\hat{\tau}+1)^2 - 4\hat{\rho}^2\hat{\tau}}$$

where $\hat{\tau} = S_y^2 / S_x^2$ and $\hat{\rho}$ is the sample correlation coefficient between x and y , one may test the null hypothesis of equal variances using the null test of zero correlation.

There are interesting repercussions to this test. For example, the power-efficiency of the Morgan-Pitman test for paired data is generally ($\rho \neq 0$) far greater than the usual F-test for independent samples (Kraemer, 1981).

Furthermore, one can obtain a joint confidence interval for the variance ratio and the correlation between the original paired measures by obtaining a confidence interval for the correlation between the pair sums and differences.

It may not be obvious why such a confidence interval would be of any importance. Let me give an example. Suppose one had two different ways of

measuring the same characteristics, a short and long form (the short form being the first section of the long form) of a test, for example. That one of the two is the more reliable which has smaller variance. Since the short form is included in the long form, from each subject tested on the long form, we automatically have both the short and long form scores: paired data. The Morgan-Pitman test can be used to test whether the variances substantially differ and hence the reliabilities. Here the correlation between paired scores is the geometric mean of the two test reliability coefficients. The variance ratio is the ratio of the two test reliability coefficients. Then the joint confidence interval on the pairwise correlation and variance ratio can be reformulated as a joint confidence interval on the two test reliability coefficients. In what may seem like spectacular statistical slight-of-hand, never in this procedure was test-retest or interobserver data on either the short or long form of the test actually collected. Yet the procedure yields both tests and confidence intervals.

Now some variations on this theme. If one obtained data in groups, i.e. n_x repeated observations of x and n_y repeated observations of y from each of N subjects, what then? In this case, one could summarize the data for each subject by a weighted sum and difference of the x -means and y -means:

$$\begin{aligned} w_x \bar{x} + w_y \bar{y} , \\ w_x \bar{x} - w_y \bar{y} . \end{aligned}$$

The weights are based on the number of observations of each variable per subject and the reliability of its measurement:

$$\frac{1}{w_x^2} = \frac{1 + \rho_x(n_x - 1)}{n_x} , \quad \frac{1}{w_y^2} = \frac{1 + \rho_y(n_y - 1)}{n_y} .$$

The correlation coefficient between these weighted sums and differences is zero if and only if the variances of x and y are equal. Again one might use correlation theory to resolve a difficult problem in comparing variances, a solution far more powerful than standard methods.

And if data were only partially paired? It has been demonstrated (Kraemer, 1983) that one may use the correlation coefficient of the paired portion of the sample to estimate missing paired values and then use the Morgan-Pitman test with reduced degrees of freedom. Again the resulting test is more powerful than any other test proposed to date (Ekbohm, 1982).

To complete the picture, it should be noted that Kristof (1982) has shown that to detect heterogeneity among a larger number of variances, one can correlate strategically selected linear combinations of the original variables. Once again, detecting heterogeneity of variance by applying correlation theory to matched sets of data is more powerful than standard methods, requiring independent samples.

It should not be thought that having a greater breadth and depth to one's knowledge of correlation theory is a technician's tool, only useful to devising new statistical instrumentation. The last example, accordingly, has to do with using correlation theory not to analyse, but to interpret the outcome of an analysis.

To Interpret Profiles Analysis of Variance

Test profiles are in common use (SAT, MCAT, MMPI, etc.) and very useful indeed. This has provided the inspiration to use profiles to define or to compare groups of subjects otherwise difficult to specify: autistic children, schizophrenics, etc. In this context, the items are

not test scores, well and carefully constructed and documented as to reliability and validity, and limited in number, but a list of scores, frequently very long, and having at best face validity: Item Profiles as opposed to Test Profiles.

To test whether there is homogeneity of Item Profiles within a group or to test contrasts between groups, Analysis of Variance (ANOVA) has been proposed (Bartko, 1980) (Greenhouse and Geisser, 1959). For example, suppose we had two replicate M-item profiles on each of N subjects. Here the suggested ANOVA is the simple two-way layout, with two replicates per cell. One obtains an F-test for Subjects, one for Items and one for Subjects x Items.

If the F-test for subjects is significant, what this is taken to mean is that subjects are not homogeneous. In the ANOVA model, however, it could seem that what this also means is that the pairwise correlation between items on the profile (assumed to be equal) is large. (To those familiar with the relationship of intraclass correlation and ANOVA, this is no surprise.) As a result, if one includes many redundant items in the profile, items essentially measuring the same thing, one is likely to conclude that subjects are heterogeneous. On the other hand, if the items on the profile are orthogonal, one is likely to conclude the subjects are homogeneous.

Since the profiles of which I speak tend to contain a great deal of redundancy, they tend to be judged inadequate to identify homogeneous groups of subjects (cf. Bartko, 1980).

Furthermore, it is recommended that the F-test for items simply be ignored (Bartko, 1980). One would know 'a priori', it is argued, that item means are not homogeneous. However, the F-test for items is equivalent to a test of correlation between paired subjects' responses. Correlation between paired subjects' responses is a measure of the similarity of "shape" of two profiles. Frequently, the point of a study of profiles

is to demonstrate that within a group the shapes of profiles are similar, and that there are differences in shape between groups. Yet the very test specific to this question is discarded!

Finally the F-test for Subjects x Items tests for non-parallel profiles. However, this test also is sensitive to unreliability of item scores. One may obtain a significant Subject x Item effect if and only if items are reliably measured. One would think then that a significant interaction effect would be a cause for rejoicing. On the contrary, it is interpreted as a sign of heterogeneity of shapes. Thus if one measures items well, one will tend to discard the profile.

As it currently stands, if unreliable, irrelevant, and redundant items are included in a profile, then there will be no evidence of within group heterogeneity or nonparallel shapes, and the profile will be judged "good". If items are reliable, and tap orthogonal characteristics, heterogeneity of subjects and nonparallel profiles are likely to be found, and the profiles will be judged "bad". The difficulty lies, not in the ANOVA, but in the selection of items and in the interpretation of results. Before ANOVA is used, unreliable and irrelevant items should be discarded, and redundant items combined or deleted. Then, a non-significant F-test for subjects and a significant F-test for interaction should be expected. It is the hitherto discarded F-test for items which is the crucial factor in determining the quality of the profile (Kraemer, 1983).

Conclusion

From a course on ANOVA, one gains not simply a set of techniques, but a way of thinking about data, of understanding what inferences one may or may not draw from data about populations. Correlation theory too,

comprises not just an interpretable measure of bivariate association, and a test of its significance. First of all, there is a body of many other techniques frequently untouched in training. Most important, however, is the fact that it represents yet another, complementary, way of thinking, outlook or philosophy, that can be enormously useful both to the statistician and to the researcher.

References

- Bartko, J.J., Kramer, M., Williams, K.A. Multivariate Methods for Comparing Patient Profiles. A Report from the International Pilot Study of Schizophrenia. National Institute of Mental Health, Series GN No. 1, 1980.
- Ekbohm, G. On Comparing Variances in the Paired Case with Incomplete Data. Biometrika, 1982, 69, 670-673.
- Glass, G.V., McGaw, B. and Smith, M.L. Meta-Analysis in Social Research. Beverly Hills: Sage, 1991.
- Greenhouse, S. and Geisser, S. On Methods in the Analysis of Profile Data. Psychometrika, 1959, 24, 95-112.
- Hedges, L.V. Distribution Theory for Glass's Estimation of Effect Size and Related Estimators. Journal of Educational Statistics, 1981, 6, 107-128.
- Hedges, L.V. and Olkin, I. Clustering Estimates of Effect Magnitude from Independent Studies. Technical Report No. 173, April, 1981. Department of Statistics, Stanford University, Stanford, CA.
- Kraemer, H.C. Robustness of the Distribution Theory of the Product Moment Correlation Coefficient. Journal of Educational Statistics, 1980, 5, 1-128.
- Kraemer, H.C. Extension of Feldt's Approach to Testing Homogeneity of Coefficients of Reliability. Psychometrika, 1981, 46, 41-45.
- Kraemer, H.C. Theory of Estimation and Testing of Effect Sizes: Use in Meta-Analysis. Journal of Educational Statistics, in press.
- Kraemer, H.C. Reconsideration of the Problem of Comparing Variances with Partially Paired Data. In preparation.
- Kristof, W. Contributions to the Analysis of Correlation Coefficients. Ph.D. Dissertation, University of South Africa, 1982.
- Moore, M. The Use of Statistical Techniques by Psychological Journal Articles. The American Statistician, 1975, 29, 68-69.
- Morgan, W.A. A Test for the Significance of the Difference between the Two Variances in a Sample from a Normal Bivariate Population. Piometrika, 1939, 31, 13-19.
- Pitman, E.J.G. A Note on Normal Correlation. Biometrika, 1939, 31, 9-12.